

# A Martingale Proof of Dobrushin's Theorem for Non-Homogeneous Markov Chains

S. Sethuraman and S.R.S. Varadhan  
Iowa State University and New York University

February 1, 2008

## Abstract

In 1956, Dobrushin proved a definitive central limit theorem for non-homogeneous Markov chains. In this note, a shorter and different proof elucidating more the assumptions is given through martingale approximation.

Partially supported by NSF/DMS-0071504 and NSF/DMS-0104343.

E-mail: sethuran@iastate.edu and varadhan@cims.nyu.edu

*Key words and phrases:* non-homogeneous Markov, contraction coefficient, central limit theorem, martingale approximation

*Abbreviated title:* Martingale Proof of Dobrushin's Theorem

*AMS (2000) subject classifications:* Primary 60J10; secondary 60F05.

## 1 Introduction and Results

Nearly fifty years ago, R. Dobrushin proved in his thesis [2] a definitive central limit theorem (CLT) for Markov chains in discrete time that are not necessarily homogeneous in time. Previously, Markov, Bernstein, Sapagov, and Linnik, among others, had considered the central limit question under various sufficient conditions. Roughly, the progression of results relaxed the state space structure from 2 states to an arbitrary set of states, and also the level of asymptotic degeneracy allowed for the transition probabilities of the chain.

After Dobrushin's work, some refinements and extensions of his CLT, some of which under more stringent assumptions, were proved by Statulevicius [16] and Sarymsakov [13]. See also Hanen [6] in this regard. A corresponding invariance principle was also proved by Gudinas [4]. More general references on non-homogeneous Markov processes can be found in Isaacson and Madsen [7], Iosifescu [8], Iosifescu and Theodorescu [9], and Winkler [18].

We now define what is meant by “degeneracy.” Although there are many measures of “degeneracy,” the measure which turns out to be most useful to work with is that in terms of the contraction coefficient. This coefficient has appeared in early results concerning Markov chains, however, in his thesis, Dobrushin popularized its use, and developed many of its important properties. [See Seneta [14] for some history.]

Let  $\pi = \pi(x, dy)$  be a Markov transition probability on  $(\mathbf{X}, \mathcal{B}(\mathbf{X}))$ . Define the contraction coefficient  $\delta(\pi)$  of  $\pi$  as

$$\begin{aligned}\delta(\pi) &= \sup_{x_1, x_2 \in \mathbf{X}, A \in \mathcal{B}} |\pi(x_1, A) - \pi(x_2, A)| \\ &= \frac{1}{2} \sup_{f: |f| \leq 1} \left| \int_{\mathbf{X}} f(y) [\pi(x_1, dy) - \pi(x_2, dy)] \right|\end{aligned}$$

Also, define the related coefficient  $\alpha(\pi) = 1 - \delta(\pi)$ .

Clearly,  $0 \leq \delta(\pi) \leq 1$ , and  $\delta(\pi) = 0$  if and only if  $\pi(x, dy)$  is independent of  $x$ . It makes sense to call  $\pi$  “non-degenerate” if  $0 \leq \delta(\pi) < 1$ . We use the standard convention and denote by  $\mu\pi$  and  $\pi u$  the transformations induced by  $\pi$  on countably additive signed measures and bounded measurable functions respectively,

$$(\mu\pi)(A) = \int \pi(x, A) \mu(dx) \quad \text{and} \quad (\pi u)(x) = \int u(y) \pi(x, dy)$$

It is easy to see that  $\delta(\pi)$  has the following properties.

$$\delta(\pi) = \sup_{\substack{x_1, x_2 \\ u \in \mathcal{U}}} |(\pi u)(x_1) - (\pi u)(x_2)|$$

where  $\mathcal{U} = \{u : \sup_{y_1, y_2} |u(y_1) - u(y_2)| \leq 1\}$ . It is the operator norm of  $\pi$  with respect to the Banach (semi-) norm  $\text{Osc}(u) = \sup_{x_1, x_2} |u(x_1) - u(x_2)|$ , namely the oscillation of  $u$ . In particular, for any transition probabilities  $\pi_1, \pi_2$  we have

$$\delta(\pi_1 \pi_2) \leq \delta(\pi_1) \delta(\pi_2). \quad (1.1)$$

If  $\mu$  is a signed measure with  $\mu(\mathbf{X}) = 0$ ,

$$\|\mu\|_{\text{Var}} = \sup_A |\mu(A)| = \frac{1}{2} \sup_{\|u\|_{L^\infty} \leq 1} \left| \int f(x) \mu(dx) \right| = \sup_{u \in \mathcal{U}} \left| \int u(x) \mu(dx) \right|.$$

Therefore, by duality, for any two probability measures  $\lambda$  and  $\mu$  on  $\mathbf{X}$ ,

$$\|(\lambda - \mu)\pi\|_{\text{Var}} \leq \delta(\pi) \|\lambda - \mu\|_{\text{Var}}. \quad (1.2)$$

By a non-homogeneous Markov chain of length  $n$  on state space  $(\mathbf{X}, \mathcal{B}(\mathbf{X}))$  corresponding to transition operators  $\{\pi_{i,i+1} = \pi_{i,i+1}(x, dy) : 1 \leq i \leq n-1\}$  we mean the Markov process  $P$  on the product space  $(\mathbf{X}^n, \mathcal{B}(\mathbf{X}^n))$ ,

$$P[X_{i+1} \in A | X_i = x] = \pi_{i,i+1}(x, A),$$

where  $\{X_i : 1 \leq i \leq n\}$  are the canonical projections. In particular, under the initial distribution  $X_1 \sim \mu$ , the distribution at time  $k \geq 1$  is  $\mu\pi_{1,2}\pi_{2,3} \cdots \pi_{k-1,k}$ . For  $i < j$  we will define

$$\pi_{i,j} = \pi_{i,i+1}\pi_{i+1,i+2} \cdots \pi_{j-1,j}$$

We denote by  $E[Z]$  and  $V(Z)$  the expectation and variance of the random variable  $Z$  with respect to  $P$ .

Consider now a non-homogeneous Markov chain on  $\mathbf{X}$  with respect to transition operators  $\{\pi_{i,i+1} : 1 \leq i \leq n-1\}$ . The following comparison of marginal distributions at time  $n$  starting from different initial conditions is an easy consequence of (1.1) and (1.2).

$$\begin{aligned} \|\lambda\pi_{1,n} - \mu\pi_{1,n}\|_{\text{Var}} &\leq \|\mu - \nu\|_{\text{Var}} \delta(\pi_{1,n}) \\ &\leq \|\mu - \nu\|_{\text{Var}} \prod_{i=1}^{n-1} \delta(\pi_{i,i+1}). \end{aligned} \quad (1.3)$$

Dobrushin's theorem concerns the fluctuations of an array of non-homogeneous Markov chains. For each  $n \geq 1$ , let  $\{X_i^{(n)} : 1 \leq i \leq n\}$  be  $n$  observations of a non-homogeneous Markov chain on  $\mathbf{X}$  with transition matrices  $\{\pi_{i,i+1}^{(n)} = \pi_{i,i+1}^{(n)}(x, dy) : 1 \leq i \leq n-1\}$ . Let also

$$\alpha_n = \min_{1 \leq i \leq n-1} \alpha(\pi_{i,i+1}^{(n)}).$$

In addition, let  $\{f_i^{(n)} : 1 \leq i \leq n\}$  be real valued functions on  $\mathbf{X}$ . Define, for  $n \geq 1$ , the sum

$$S_n = \sum_{i=1}^n f_i^{(n)}(X_i^{(n)}).$$

**Theorem 1.1** *Suppose that for some finite constants  $C_n$ ,*

$$\sup_x \sup_{1 \leq i \leq n} |f_i^{(n)}(x)| \leq C_n.$$

*Then, if*

$$\lim_{n \rightarrow \infty} C_n^2 \alpha_n^{-3} \left[ \sum_{i=1}^n V(f_i^{(n)}(X_i^{(n)})) \right]^{-1} = 0, \quad (1.4)$$

*we have, regardless of the initial distribution, that*

$$\frac{S_n - E[S_n]}{\sqrt{V(S_n)}} \Rightarrow N(0, 1). \quad (1.5)$$

*In general, the result is not true if condition (1.4) is not met.*

In [2], Dobrushin also states the direct corollary which simplifies some of the assumptions.

**Corollary 1.1** *When the functions are uniformly bounded, i.e.  $\sup_n C_n = C < \infty$  and the variances are bounded below, i.e.  $V(f_i^{(n)}(X_i^{(n)})) \geq c > 0$ , for all  $1 \leq i \leq n$  and  $n \geq 1$ , then we have the convergence (1.5) provided*

$$\lim_{n \rightarrow \infty} n^{1/3} \alpha_n = \infty. \quad (1.6)$$

We remark that in [2] (e.g. Theorems 3, 8) there are also results where the boundedness condition on  $f_i^{(n)}$  is replaced by integrability conditions. As these results follow from truncation methods and Theorem 1.1 for bounded variables, we only consider Dobrushin's theorem in the bounded case.

Also, for the ease of the reader, and to be complete, we will discuss in the next section an example, given in [2] and due to Dobrushin and Bernstein, of how the weak convergence (1.5) may fail when the condition (1.4) is not satisfied.

We now consider Dobrushin's methods. The techniques used in [2] to prove the above results fall under the general heading of the "blocking method." The condition (1.4) ensures that well-separated blocks of observations may be approximated by independent versions with small error. Indeed, in many remarkable steps, Dobrushin exploits the Markov property and several contraction coefficient properties, which he himself derives, to deduce error bounds sufficient to apply CLT's for independent variables. However, in [2], it is difficult to see, even at the technical level, why condition (1.4) is natural.

The aim of this note is to provide a different, shorter proof of Theorem 1.1 which explains more why condition (1.4) appears in the result. The methods are through martingale approximations and martingale CLT's which perhaps were not as codified in the early 1950's as they are today. These methods go back at least to Gordin [3] in the context of homogeneous processes, and has been used by others in other "related" situations (e.g. Kifer [10]; see also Pinsky [12]). There are three main ingredients in this approximation with respect to the non-homogeneous setting of Theorem 1.1, (1) negligibility estimates for individual components, (2) a law of large numbers for conditional variances, and (3) lower bounds for the variance  $V(S_n)$ . Negligibility bounds and a LLN are well known requirements for martingale CLT's (cf. Hall-Heyde [5, ch. 3]), and in fact, as will be seen, the sufficiency of condition (1.4) is transparent in the proofs of these two components (Lemma 4.2, and Lemmas 4.3 and 4.4). The variance lower bounds which we will use were as well derived by Dobrushin in his proof. However, using some martingale properties, we give a more direct argument for a better estimate.

We note also, with this martingale approximation, that an invariance principle for the partial sums holds through standard martingale propositions, Hall-Heyde [5], among other results. In fact, from the martingale invariance principle, it should be possible to derive Gudynas's theorems [4] although this is not done here.

We now explain the structure of the article. In section 2, we give the Bernstein-Dobrushin example of a Markov chain with anomalous behavior. In section 3, we discuss needed properties of the contraction coefficient. In section 4, we state the martingale CLT that will be utilized, and, as a preview of the non-homogeneous chain proof, we quickly reprise the argument with respect to homogeneous chains. In section 5, we prove Theorem 1.1 with martingale approximation assuming a lower bound on the variance  $V(S_n)$ . And last, in section 6, we prove this variance estimate.

## 2 Anomalous Example

Here, we summarize the example in Dobrushin's thesis, attributed to Bernstein, which shows that condition (1.4) is sharp.

**Example 2.1** Let  $\mathbf{X} = \{1, 2\}$ , and consider the  $2 \times 2$  transition matrices on  $\mathbf{X}$ ,

$$Q(p) = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}$$

The contraction coefficient  $\delta(Q(p))$  of  $Q(p)$  is  $|1-2p|$ . Note that  $\delta(Q(p)) = \delta(Q(1-p))$ . The invariant measures for all the  $Q(p)$  are the same  $p(1) = p(2) = \frac{1}{2}$ . We will be looking at  $Q(p)$  for  $p$  close to 0 or 1 and the special case of  $p = \frac{1}{2}$ . However, when  $p$  is small, the homogeneous chains behave very differently under  $Q(p)$  and  $Q(1-p)$ . More specifically, when  $p$  is small there are very few switches between the two states whereas when  $1-p$  is small it switches most of the time. In fact, this behavior can be made more precise (see Dobrushin [1], Hanen [6], or from direct computation). Let  $T_n = \sum_{i=1}^n \mathbf{1}_{\{1\}}(X_i)$  count the number of visits to state 1, say, in  $n$  steps.

*Case A.* Consider the homogeneous chain under  $Q(p)$  with  $p = \frac{1}{n}$  and initial distribution  $p(1) = p(2) = \frac{1}{2}$ . Then,

$$\frac{T_n}{n} \Rightarrow G \quad \text{and} \quad \lim_{n \rightarrow \infty} n^{-2} V(T_n) = V_A < \infty. \quad (2.1)$$

where  $G$  is a proper distribution supported on  $[0, 1]$ .

*Case B.* Consider the homogeneous chain run under  $Q(p)$  with  $p = 1 - \frac{1}{n}$  and initial distribution  $p(1) = p(2) = \frac{1}{2}$ . Then,

$$T_n - \frac{n+1}{2} \Rightarrow F \quad \text{and} \quad \lim_{n \rightarrow \infty} V(T_n) = V_B < \infty. \quad (2.2)$$

where  $F$  is a proper distribution function.

Let a sequence  $\alpha_n \rightarrow 0$  with  $\alpha_n \geq n^{-\frac{1}{3}}$  be given. To construct the anomalous Markov chain, it will be helpful to split the time horizon  $[1, 2, \dots, n]$  into roughly  $n\alpha_n$  blocks of size  $\alpha_n^{-1}$ . We interpose a  $Q(\frac{1}{2})$  between any two blocks that has the effect

of making the blocks independent of each other. More precisely let  $k_i^{(n)} = i[\alpha_n^{-1}]$  for  $1 \leq i \leq m_n$  where  $m_n = [n/\alpha_n^{-1}]$ . Also, define  $k_0^{(n)} = 0$ , and  $k_{m_n+1}^{(n)} = n$ .

Define now, for  $1 \leq i \leq n$ ,

$$\pi_{i,i+1}^{(n)} = \begin{cases} Q(\alpha_n) & \text{for } i = 1, 2, \dots, k_1^{(n)} - 1 \\ Q(\frac{1}{2}) & \text{for } i = k_1^{(n)}, k_2^{(n)}, \dots, k_{m_n}^{(n)} \\ Q(1 - \alpha_n) & \text{for all other } i. \end{cases}$$

Consider the non-homogeneous chain with respect to  $\{\pi_{i,i+1}^{(n)} : 1 \leq i \leq n-1\}$  starting from equilibrium  $p(0) = p(1) = \frac{1}{2}$ . From the definition of the chain, one observes, as  $Q(\frac{1}{2})$  does not distinguish between states, that the process in time horizons  $\{(k_i^{(n)} + 1, k_{i+1}^{(n)}) : 0 \leq i \leq m_n\}$  are mutually independent. For the first time segment 1 to  $k_1^{(n)}$ , the chain is in regime  $A$ , while for the other segments, the chain is in case  $B$ .

Once again, let us concentrate on the number of visits to state 1. Denote by  $T^{(n)} = \sum_{i=1}^n \mathbf{1}_{\{1\}}(X_i^{(n)})$  and  $T^{(n)}(k, l) = \sum_{i=k}^l \mathbf{1}_{\{1\}}(X_i^{(n)})$  the counts in the first  $n$  steps and in steps  $k$  to  $l$  respectively. It follows from the discussion of independence above that

$$T^{(n)} = \sum_{i=0}^{m_n} T^{(n)}(k_i^{(n)} + 1, k_{i+1}^{(n)})$$

is the sum of independent sub-counts where, additionally, the sub-counts for  $1 \leq i \leq m_n - 1$  are identically distributed, the last sub-count perhaps being shorter. Also, as the initial distribution is invariant, we have  $V(\mathbf{1}_{\{1\}}(X_i^{(n)})) = 1/4$  for all  $i$  and  $n$ . Then, in the notation of Corollary 1.1,  $C = 1$  and  $c = 1/4$ .

From (2.1), we have that

$$V(T^{(n)}(1, k_1^{(n)})) \sim \alpha_n^{-2} V_A \text{ as } n \uparrow \infty.$$

Also, from (2.2) and independence of  $m_n$  sub-counts, we have that

$$V(T^{(n)}(k_1^{(n)} + 1, n)) \sim n \alpha_n V_B \text{ as } n \uparrow \infty.$$

From these calculations, we see if  $n^{1/3} \alpha_n \rightarrow \infty$ , then  $\alpha_n^{-2} \ll n \alpha_n$ , and so the major contribution to  $T^{(n)}$  is from  $T^{(n)}(k_1^{(n)} + 1, n)$ . However, since this last count is (virtually) the sum of  $m_n$  i.i.d. sub-counts, we have that  $T^{(n)}$ , properly normalized, converges to  $N(0, 1)$ , as predicted by Dobrushin's theorem 1.1.

On the other hand, if  $\alpha_n = n^{-1/3}$ , we have  $\alpha_n^{-2} = n \alpha_n$ , and count  $T^{(n)}(1, k_1^{(n)})$ , independent of  $T^{(n)}(k_1^{(n)}, n)$ , also contributes to the sum  $T^{(n)}$ . After appropriate scaling, then,  $T^{(n)}$  approaches the convolution of a non-trivial non-normal distribution and a normal distribution, and therefore is certainly not Gaussian.

### 3 Martingale CLT

The central limit theorem for martingale differences is by now a standard tool. We quote the following (strong) form of the result implied by Corollary 3.1 in Hall and Heyde [5].

**Proposition 3.1** *For each  $n \geq 1$ , let  $\{(W_i^{(n)}, \mathcal{F}_i^{(n)}) : 0 \leq i \leq n\}$  be a martingale relative to the nested family  $\mathcal{F}_i^{(n)} \subset \mathcal{F}_{i+1}^{(n)}$  with  $W_0^{(n)} = 0$ . Let  $\xi_i^{(n)} = W_i^{(n)} - W_{i-1}^{(n)}$  be their differences. Suppose that*

$$\begin{aligned} \max_{1 \leq i \leq n} \|\xi_i^{(n)}\|_{L^\infty} &\rightarrow 0 & \text{and} \\ \sum_{i=1}^n E[(\xi_i^{(n)})^2 | \mathcal{F}_{i-1}^{(n)}] &\rightarrow 1 & \text{in } L^2. \end{aligned}$$

Then,

$$W_n^{(n)} \Rightarrow N(0, 1).$$

Note that the first and second limit conditions are the negligibility assumption on the sequence, and law of large numbers for conditional variances mentioned in the introduction.

We now sketch a proof of Corollary 1.1 in the case of a homogeneous Markov chain on a finite state space. Assume that we have a Markov chain with transition probability  $P$  on a finite state space  $\mathbf{X}$ . If  $\delta(P) < 1$ , and  $f : \mathbf{X} \rightarrow R$  is a function with mean 0 with respect to the invariant distribution  $\pi$  on  $\mathbf{X}$ , it is in the range of  $I - P$  and the equation  $(I - P)u = f$  has a solution. The following argument is implicit in Gordin [3], and also explicitly used in Kipnis and Varadhan [11].

Using the relation  $E[u(X_{j+1}) | \mathcal{F}_j] = (Pu)(X_j)$ , it is easy to check that

$$f(X_j) = u(X_j) - E[u(X_{j+1}) | \mathcal{F}_j] = u(X_j) - u(X_{j+1}) + \xi_{j+1}$$

where

$$\xi_j = u(X_j) - E[u(X_j) | \mathcal{F}_{j-1}]$$

is a martingale difference. Then,

$$\sum_{j=0}^{n-1} f(X_j) = u(X_0) - u(X_n) + \sum_{j=1}^n \xi_j.$$

If we define

$$E[\xi_{j+1}^2 | \mathcal{F}_j] = q(X_j)$$

We will apply the martingale CLT (Proposition 3.1) to the array formed from  $W_i^{(n)} = M_i/\sqrt{n}$  with differences  $\xi_i^{(n)} = \xi_i/\sqrt{n}$ . As the differences are uniformly bounded,

$\|\xi_i^{(n)}\|_{L^\infty} \leq 2\|u\|_{L^\infty}/\sqrt{n}$ , the first condition of Proposition 3.1 is satisfied. The second follows from the following computation. From the Markov property,

$$\sum_{i=1}^n E[(\xi_i^{(n)})^2 | \mathcal{F}_{i-1}^{(n)}] = \frac{1}{n} \sum_{i=1}^n E[\xi_i^2 | X_{i-1}] = \frac{1}{n} \sum_{i=0}^{n-1} q(X_i).$$

So, by the ergodic theorem, the last expression converges almost surely to  $V_0 = E_\pi[q(X_0)] < \infty$ . It is not difficult to see that  $V_0 > 0$ .

Therefore,  $V(M_n) \sim nV_0$  and  $(nV_0)^{-1/2}M_n \Rightarrow N(0, 1)$  by Proposition 3.1. Since the difference

$$n^{-\frac{1}{2}}|M_n - \sum_{j=0}^{n-1} f(X_j)| \leq n^{-1/2}\|u(X_n) - u(X_0)\|_{L^\infty} \rightarrow 0$$

we have  $V(S_n) \sim nV_0$  and  $S_n/\sqrt{V(S_n)} \Rightarrow N(0, 1)$  also.  $\square$

## 4 Proof of Theorem 1.1

We give here a short proof for Theorem 1.1 through martingale approximation, illustrated for homogeneous chains in the previous section. Consider the non-homogeneous setting of Theorem 1.1. To follow the homogeneous argument, we will need to find the non-homogeneous analogue of the resolvent function “ $u = (I - P)^{-1}f$ .” To simplify notation, we will assume throughout that the functions  $\{f_i^{(n)}\}$  are mean-zero,  $E[f_i^{(n)}(X_i^{(n)})] = 0$  for  $1 \leq i \leq n$  and  $n \geq 1$ . Define

$$Z_k^{(n)} = \sum_{i=k}^n E[f_i^{(n)}(X_i^{(n)}) | X_k^{(n)}]$$

where

$$Z_k^{(n)} = \begin{cases} f_k^{(n)}(X_k^{(n)}) + \sum_{i=k+1}^n E[f_i^{(n)}(X_i^{(n)}) | X_k^{(n)}] & \text{for } 1 \leq k \leq n-1 \\ f_k^{(n)}(X_n^{(n)}) & \text{for } k = n. \end{cases} \quad (4.1)$$

**Remark 4.1** Before going further, we remark that indeed sequence  $\{Z_k^{(n)}\}$  can be thought of as a generalization of the resolvent sequence  $\{u(X_k)\}$  used in the case of a homogeneous chain. When the array  $\{X_i^{(n)}\}$  is formed from the sequence  $\{X_i\}$ ,  $f_i^{(n)} = f$  for all  $i$  and  $n$ , and the chain is homogeneous,  $P_n = P$  for all  $n$ , then indeed  $Z_k^{(n)}$  reduces to  $Z_k^{(n)} = f(X_k) + \sum_{i=1}^{n-k} (P^i f)(X_k)$  which approximates  $\sum_{i=0}^{\infty} (P^i f)(X_k) = [(I - P)^{-1}f](X_k) = u(X_k)$ . See also p. 145-6 Varadhan [17] for other uses of  $\{Z_k^{(n)}\}$ .



Now, let us return to the full non-homogeneous setting of Theorem 1.1. By rearranging terms in (4.1), we obtain for  $1 \leq k \leq n-1$

$$\begin{aligned} f_k^{(n)}(X_k^{(n)}) &= Z_k^{(n)} - E[Z_{k+1}^{(n)} | X_k^{(n)}] \\ &= [Z_k^{(n)} - E[Z_k^{(n)} | X_{k-1}^{(n)}]] + [E[Z_k^{(n)} | X_{k-1}^{(n)}] - E[Z_{k+1}^{(n)} | X_k^{(n)}]]. \end{aligned} \quad (4.2)$$

Then, we have the decomposition,

$$\begin{aligned} S_n &= \sum_{k=1}^n f_k^{(n)}(X_k^{(n)}) \\ &= \sum_{k=2}^n [Z_k^{(n)} - E[Z_k^{(n)} | X_{k-1}^{(n)}]] + Z_1^{(n)} \end{aligned} \quad (4.3)$$

and so in particular  $V(S_n) = \sum_{k=2}^n V(Z_k^{(n)} - E[Z_k^{(n)} | X_{k-1}^{(n)}]) + V(Z_1^{(n)})$ . Let us now define the differences

$$\xi_k^{(n)} = \frac{1}{\sqrt{V(S_n)}} [Z_k^{(n)} - E[Z_k^{(n)} | X_{k-1}^{(n)}]] \quad (4.4)$$

and the martingale  $M_k^{(n)} = \sum_{l=2}^k \xi_l^{(n)}$  with respect to  $\mathcal{F}_k^{(n)} = \sigma\{X_l^{(n)} : 1 \leq l \leq k\}$  for  $n \geq 1$ . The plan to obtain Theorem 1.1 will now be to approximate  $S_n$  by  $M_n^{(n)}$  and use Proposition 3.1. Condition (1.4) will be a natural sufficient condition for “negligibility” (Lemma 4.2) and “LLN” (Lemmas 4.3 and 4.4) with regard to Proposition 3.1.

**Lemma 4.1** *For  $1 \leq i \leq j \leq n$ , we have the bound*

$$\|E[f_j^{(n)}(X_j^{(n)}) | X_i^{(n)}]\|_{L^\infty} \leq 2C_n(1 - \alpha_n)^{j-i}.$$

Hence, for  $1 \leq k \leq n$ ,

$$\|Z_k^{(n)}\|_{L^\infty} \leq 2C_n \alpha_n^{-1}.$$

*Proof.* Since  $\|f_j^{(n)}\|_{L^\infty} \leq C_n$  its oscillation  $\text{Osc}(f_j^{(n)}) \leq 2C_n$ . From (1.3),

$$\text{Osc}(\pi_{i,j} f_j^{(n)}) \leq 2C_n \delta(\pi_{i,j}) \leq 2C_n(1 - \alpha_n)^{j-i}$$

Because  $E[(\pi_{i,j} f_j^{(n)})(X_i)] = E[f_j^{(n)}(X_j)] = 0$ ,

$$\|\pi_{i,j} f_j^{(n)}\|_{L^\infty} \leq \text{Osc}(\pi_{i,j} f_j^{(n)}) \leq 2C_n(1 - \alpha_n)^{j-i}$$

The second estimate now follows from this estimate. Indeed,

$$|Z_k^{(n)}| = \left| \sum_{i=k}^n [E[f_i^{(n)}(X_i^{(n)}) | X_k^{(n)}]] \right| \leq 2C_n \sum_{i=k}^n (1 - \alpha_n)^{i-k} \leq 2C_n \alpha_n^{-1}. \quad \square$$

We now state a lower bound for the variance which will be proved in the next section using martingale ideas. We remark in [2] that actually the bound,  $V(S_n) \geq (\alpha_n/8) \sum_{i=1}^n V(f_i^{(n)}(X_i^{(n)}))$ , is found by different methods (see also section 1.2.2 [9]).

**Proposition 4.1** For  $n \geq 1$ ,

$$V(S_n) \geq \frac{\alpha_n}{4} \sum_{i=1}^n V(f_i^{(n)}(X_i^{(n)})). \quad (4.5)$$

The next estimate shows that the asymptotics of  $S_n/\sqrt{V(S_n)}$  depend only on the martingale approximant  $M_n^{(n)}$ , and that the differences  $\xi_k^{(n)}$  are negligible.

**Lemma 4.2** Under condition (1.4), we have that

$$\lim_{n \rightarrow \infty} \sup_{1 \leq k \leq n} \frac{\|Z_k^{(n)}\|_{L^\infty}}{\sqrt{V(S_n)}} = 0.$$

*Proof.* By Lemma 4.1 and Proposition 4.1,

$$\frac{\|Z_k^{(n)}\|_{L^\infty}}{\sqrt{V(S_n)}} \leq \frac{4C_n}{(\alpha_n^3 \sum_{i=1}^n V(f_i^{(n)}(X_i^{(n)})))^{1/2}}.$$

The lemma follows now from (1.4).  $\square$

We now turn to showing the LLN part of Proposition 3.1 for the array  $\{M_k^{(n)}\}$ .

**Lemma 4.3** Let  $\{Y_l^{(n)} : 1 \leq l \leq n\}$  and  $\{\mathcal{F}_l^{(n)} : 1 \leq l \leq n\}$ , for  $n \geq 1$ , be respectively an array of non-negative variables and  $\sigma$ -fields such that  $\sigma\{Y_1^{(n)}, \dots, Y_l^{(n)}\} \subset \mathcal{F}_l^{(n)}$ . Suppose that

$$\lim_{n \rightarrow \infty} E\left[\sum_{l=1}^n Y_l^{(n)}\right] = 1 \quad \text{and} \quad \sup_{1 \leq i \leq n} \|Y_i^{(n)}\|_{L^\infty} \leq \epsilon_n$$

where  $\lim_{n \rightarrow \infty} \epsilon_n = 0$ . In addition, assume

$$\lim_{n \rightarrow \infty} \sup_{1 \leq l \leq n-1} \text{Osc } E\left[\sum_{j=l+1}^n Y_j^{(n)} | \mathcal{F}_l^{(n)}\right] = 0.$$

Then,

$$\lim_{n \rightarrow \infty} \sum_{l=1}^n Y_l^{(n)} = 1 \quad \text{in } L^2.$$

*Proof.* Write

$$E\left[\left(\sum_{l=1}^n Y_l^{(n)}\right)^2\right] = \sum_{l=1}^n E[(Y_l^{(n)})^2] + 2 \sum_{l=1}^{n-1} E\left[Y_l^{(n)} \left(\sum_{j=l+1}^n Y_j^{(n)}\right)\right].$$

The first sum on the right-hand side is bounded as follows. From non-negativity,

$$\sum_{l=1}^n E[(Y_l^{(n)})^2] \leq \epsilon_n \sum_{l=1}^n E[Y_l^{(n)}] = \epsilon_n \cdot (1 + o(1)) \rightarrow 0 \quad \text{as } n \uparrow \infty.$$

Consider now the second sum. Write

$$\sum_{l=1}^{n-1} E[Y_l^{(n)} (\sum_{j=l+1}^n Y_j^{(n)})] = \sum_{l=1}^{n-1} E[Y_l^{(n)} E[\sum_{j=l+1}^n Y_j^{(n)} | \mathcal{F}_l^{(n)}]].$$

From the oscillation assumption, we have that

$$\sup_{1 \leq l \leq n-1} \sup_{\omega} |E[\sum_{j=l+1}^n Y_j^{(n)} | \mathcal{F}_l^{(n)}](\omega) - E[\sum_{j=l+1}^n Y_j^{(n)}]| = o(1).$$

Therefore,

$$\begin{aligned} 2 \sum_{l=1}^{n-1} E[Y_l^{(n)} (\sum_{j=l+1}^n Y_j^{(n)})] &= 2 \sum_{l=1}^{n-1} E[Y_l^{(n)}] E[\sum_{j=l+1}^n Y_j^{(n)}] + o(1) \cdot \sum_{l=1}^{n-1} E[Y_l^{(n)}] \\ &= (\sum_{l=1}^n E[Y_l^{(n)}])^2 - \sum_{l=1}^n E[(Y_l^{(n)})^2] + o(1) \\ &= 1 + o(1). \end{aligned}$$

Putting together these statements, we obtain the lemma.  $\square$

To apply this result to our situation, we will need the following oscillation estimate.

**Lemma 4.4** *Let  $v_l^{(n)} = E[(\xi_l^{(n)})^2 | X_{l-1}^{(n)}]$  and  $\mathcal{F}_l^{(n)} = \sigma\{X_1^{(n)}, \dots, X_j^{(n)}\}$  for  $2 \leq l \leq n$  and  $1 \leq j \leq n$ . Then, under condition (1.4), we have*

$$\sup_{2 \leq l \leq n-1} \text{Osc } E[\sum_{j=l+1}^n v_j^{(n)} | \mathcal{F}_l^{(n)}] = o(1).$$

*Proof.* From the martingale and Markov property, we have  $E[\xi_r^{(n)} \xi_s^{(n)} | X_u^{(n)}] = 0$  for  $r > s \geq u$ . Then,

$$\begin{aligned} E[\sum_{j=l+1}^n v_j^{(n)} | \mathcal{F}_l^{(n)}] &= E[\sum_{j=l+1}^n (\xi_j^{(n)})^2 | X_l^{(n)}] \\ &= E[\left(\sum_{j=l+1}^n \xi_j^{(n)}\right)^2 | X_l^{(n)}] \\ &= V(S_n)^{-1} E[(\sum_{j=l+1}^n f_j^{(n)}(X_j^{(n)}) - E[Z_{l+1}^{(n)} | X_l^{(n)}])^2 | X_l^{(n)}] \\ &= (1 + o(1)) V(S_n)^{-1} E[(\sum_{j=l+1}^n f_j^{(n)}(X_j^{(n)}))^2 | X_l^{(n)}] + o(1) \end{aligned}$$

where we rewrite  $\xi_j^{(n)}$  with (4.4) in the third line, and use Lemma 4.1 in the last line.

Therefore, let us consider oscillations of

$$E\left[\left(\sum_{j=l+1}^n f_j^{(n)}(X_j^{(n)})\right)^2 | X_l^{(n)}\right] = \sum_{j,m=l+1}^n E[f_j^{(n)}(X_j^{(n)})f_m^{(n)}(X_m^{(n)}) | X_l^{(n)}]. \quad (4.6)$$

From Lemma 4.1, we have the bound, for  $j \leq m$ ,

$$\begin{aligned} \|E[f_j^{(n)}(X_j^{(n)})f_m^{(n)}(X_m^{(n)}) | X_l^{(n)}]\|_{L^\infty} &= \|E[f_j^{(n)}(X_j^{(n)})E[f_m^{(n)}(X_m^{(n)}) | X_j^{(n)}] | X_l^{(n)}]\|_{L^\infty} \\ &\leq 4C_n^2(1 - \alpha_n)^{j-l}(1 - \alpha_n)^{m-j}. \end{aligned}$$

Therefore, the oscillations of (4.6) are bounded by  $16C_n^2\alpha_n^{-2}$  uniformly in  $l$ . Hence, using Proposition 4.1, we obtain

$$\sup_{2 \leq l \leq n-1} \text{Osc } E\left[\sum_{j=l+1}^n v_j^{(n)} | \mathcal{F}_l^{(n)}\right] \leq (16)(4)C_n^2 \left[\alpha_n^3 \sum_{j=1}^n V(f_j^{(n)}(X_j^{(n)}))\right]^{-1}$$

which is  $o(1)$  by (1.4).  $\square$

*Proof of Theorem 1.1.* From Lemma 4.2, we need only show that  $M_n^{(n)}/\sqrt{V(S_n)} \Rightarrow N(0,1)$ . This will follow from martingale convergence (Proposition 3.1) as soon as we show (1)  $\sup_{2 \leq k \leq n} \|\xi_k^{(n)}\|_{L^\infty} \rightarrow 0$  and (2)  $\sum_{k=2}^n E[(\xi_k^{(n)})^2 | \mathcal{F}_{k-1}^{(n)}] \rightarrow 1$ . However, (1) follows from the negligibility estimate Lemma 4.2, and (2) from LLN Lemmas 4.3 and 4.4 since “negligibility” (1) holds and  $\sum_{k=2}^n E[(\xi_k^{(n)})^2] = 1 + o(1)$  (from variance decomposition near (4.3) and Lemma 4.2).  $\square$

## 5 Proof of Variance Lower Bound

In this section, we prove Proposition 4.1.

**Lemma 5.1** *Let  $f$  and  $g$  be measurable functions on  $(\mathbf{X}, \mathcal{B}(\mathbf{X}))$ . Let  $\lambda$  be a probability measure on  $\mathbf{X} \times \mathbf{X}$  with marginals  $\alpha$  and  $\beta$  respectively. Let  $\pi(x_1, dx_2)$  and  $\hat{\pi}(x_2, dx_1)$  be the transition probabilities in the two directions so that*

$$\alpha\pi = \beta, \quad \beta\hat{\pi} = \alpha$$

If

$$\int f(x_1)\alpha(dx_1) = \int g(x_2)\beta(dx_2) = 0$$

then,

$$\left| \int f(x_1)g(x_2)\lambda(dx_1, dx_2) \right| \leq \sqrt{\delta(\pi)} \|f\|_{L_2(\alpha)} \|g\|_{L_2(\beta)}$$

*Proof.* Let us construct a measure on  $\mathbf{X} \times \mathbf{X} \times \mathbf{X}$  by starting with  $\lambda$  on  $\mathbf{X} \times \mathbf{X}$  and using reversed  $\widehat{\pi}(x_2, dx_3)$  to go from  $x_2$  to  $x_3$ . The transition probability from  $x_1$  to  $x_3$  defined by

$$Q(x_1, A) = \int \pi(x_1, dx_2) \widehat{\pi}(x_2, A)$$

satisfies  $\delta(Q) \leq \delta(\pi)$ . Moreover  $\alpha Q = \alpha$  and the operator  $Q$  is self adjoint and bounded with norm 1 on  $L_2(\alpha)$ . Then, if  $f$  is a bounded function with  $\int f(x) \alpha(dx) = 0$  (and so  $E_\alpha[Q^n f] = 0$ ), we have for  $n \geq 1$ ,

$$\|Q^n f\|_{L_2(\alpha)} \leq \|Q^n f\|_{L_\infty} \leq (\delta(Q))^n \text{Osc}(f). \quad (5.1)$$

Hence, as bounded functions are dense, on the subspace of functions,  $M = \{f \in L_2(\alpha) : \int f(x) \alpha(dx) = 0\}$ , the top of the spectrum of  $Q$  is less than  $\delta(Q)$  and so  $\|Q\|_{L_2(\alpha, M)} \leq \delta(Q)$ . Indeed, suppose the spectral radius of  $Q$  on  $M$  is larger than  $\delta(Q) + \epsilon$  for  $\epsilon > 0$ , and  $f \in M$  is a non-trivial bounded function whose spectral decomposition is with respect to spectral values larger than  $\delta(Q) + \epsilon$ . Then,  $\|Q^n f\|_{L_2(\alpha)} \geq \|f\|_{L_2(\alpha)} (\delta(Q) + \epsilon)^n$  which contradicts the bound (5.1) when  $n \uparrow \infty$ . [cf. Thm. 2.10 [15] for a proof in discrete space settings.]

Then,

$$\|\widehat{\pi} f\|_{L_2(\beta)}^2 = \langle \pi \widehat{\pi} f, f \rangle_{L_2(\alpha)} = \langle Q f, f \rangle_{L_2(\alpha)} \leq \|Q\|_{L_2(\alpha, M)} \|f\|_{L_2(\alpha)} \leq \delta(Q) \|f\|_{L_2(\alpha)}^2.$$

Finally

$$\left| \int f(x_1) g(x_2) \lambda(dx_1, dx_2) \right| = \left| \langle \widehat{\pi} f, g \rangle_{L_2(\beta)} \right| \leq \sqrt{\delta(\pi)} \|f\|_{L_2(\alpha)} \|g\|_{L_2(\beta)}.$$

□

**Lemma 5.2** *Let  $f(x_1)$  and  $g(x_2)$  be square integrable with respect to  $\alpha$  and  $\beta$  respectively. Then,*

$$E^\lambda \left[ (f(x_1) - g(x_2))^2 \right] \geq \alpha(\pi) V(f(\cdot))$$

as well as

$$E^\lambda \left[ (f(x_1) - g(x_2))^2 \right] \geq \alpha(\pi) V(g(\cdot))$$

*Proof.* We can assume without loss of generality that  $f$  and  $g$  have mean 0 with respect to  $\alpha$  and  $\beta$  respectively. Then

$$\begin{aligned} E^\lambda \left[ (f(x_1) - g(x_2))^2 \right] &= E^\alpha \left[ [f(x_1)]^2 \right] + E^\beta \left[ [g(x_2)]^2 \right] - 2E^\lambda \left[ f(x_1) g(x_2) \right] \\ &\geq E^\alpha \left[ [f(x_1)]^2 \right] + E^\beta \left[ [g(x_2)]^2 \right] - 2\sqrt{\delta(\pi)} \|f\|_{L_2(\alpha)} \|g\|_{L_2(\beta)} \\ &\geq (1 - \delta(\pi)) \|f\|_{L_2(\alpha)}^2 = \alpha(\pi) \|f\|_{L_2(\alpha)}^2 \end{aligned}$$

The proof of the second half is identical. □

*Proof of Proposition 4.1.* Applying Lemma 5.2 to the Markov pairs  $\{(X_k^{(n)}, X_{k+1}^{(n)}) : 1 \leq k \leq n-1\}$  with  $f(X_k^{(n)}) = E[Z_{k+1}^{(n)} | X_k^{(n)}]$  and  $g(X_{k+1}^{(n)}) = Z_{k+1}^{(n)}$ , we get

$$E[(Z_{k+1}^{(n)} - E[Z_{k+1}^{(n)} | X_k^{(n)}])^2] \geq \alpha_n E[(Z_{k+1}^{(n)})^2]$$

On the other hand from (4.2), for  $1 \leq k \leq n-1$ , we have

$$\begin{aligned} V(f_k^{(n)}(X_k^{(n)})) &\leq E[(f_k^{(n)}(X_k^{(n)}))^2] \\ &\leq 2E[(Z_k^{(n)})^2] + 2E[(E[Z_{k+1}^{(n)} | X_k^{(n)}])^2] \\ &\leq 2E[(Z_k^{(n)})^2] + 2E[(Z_{k+1}^{(n)})^2] \end{aligned}$$

Summing over  $k$ , and noting variance decomposition near (4.3),

$$\begin{aligned} \sum_{k=1}^n V(f_k^{(n)}(X_k^{(n)})) &\leq 4 \sum_{k=1}^n E[(Z_k^{(n)})^2] \\ &\leq \frac{4}{\alpha_n} \left[ \sum_{k=1}^{n-1} E[(Z_{k+1}^{(n)} - E[Z_{k+1}^{(n)} | X_k^{(n)}])^2] + E[(Z_1^{(n)})^2] \right] = \frac{4}{\alpha_n} V(S_n). \end{aligned}$$

□

## References

- [1] Dobrushin, R. (1953) Limit theorems for Markov chains with two states. (Russian) *Izv. Akad. Nauk SSSR* **17:4** 291-330.
- [2] Dobrushin, R. (1956) Central limit theorems for non-stationary Markov chains I,II. *Theory of Probab. and its Appl.* **1** 65-80, 329-383.
- [3] Gordin, M.I. (1969) The central limit theorem for stationary processes. *Soviet Math. Dokl.* **10** 1174-1176.
- [4] Gudynas, P. (1977) An invariance principle for inhomogeneous Markov chains. *Lithuanian Math. J.* **17:2** 184-192.
- [5] Hall, P. and Heyde, C.C. (1980) *Martingale limit theory and its application*. Academic Press, New York.
- [6] Hanen, A. (1963) Théorèmes limites pour une suite de chaînes de Markov. *Ann. Inst. H. Poincaré* **18** 197-301.
- [7] Isaacson, D.L., and Madsen, R.W. (1976) *Markov chains theory and applications*. John Wiley and Sons, New York.
- [8] Iosifescu, M. (1980) *Finite Markov processes and their applications*. John Wiley and Sons, New York.

- [9] Iosifescu, M., and Theodorescu, R. (1969) *Random processes and learning*. Springer, Berlin.
- [10] Kifer, Y. (1998) Limit theorems for random transformations and processes in random environments. *Trans. Amer. Math. Soc.* **350** 1481-1518.
- [11] Kipnis, C., Varadhan, S. R. S. (1986) Central limit theorem for additive functionals of reversible markov processes. *Commun. Math. Phys.* **104** 1-19.
- [12] Pinsky, M. (1991) *Lectures on Random Evolution*. World Scientific, Singapore.
- [13] Sarmysakov, T.A. (1961) Inhomogeneous Markov chains. *Theor. Probability Appl.* **6** 178-185.
- [14] Seneta, E. (1973) On the historical development of the theory of finite inhomogeneous Markov chains. *Proc. Cambridge Philos. Soc.* **74** 507-513.
- [15] Seneta, E. (1981) *Non-negative Matrices and Markov Chains*. Second Edition, Springer-Verlag, New York.
- [16] Statulevicius, V. (1969-70) Limit theorems for sums of random variables connected in Markov chains. *Liet. Mat. Rink.* **9** 346-362, 635-672, **10** 583-592.
- [17] Varadhan, S.R.S. (2001) *Probability Theory*. Courant Lecture Notes **7** American Mathematical Society, Providence, R.I.
- [18] Winkler, G. (1995) *Image analysis, random fields and dynamic Monte Carlo methods. A mathematical introduction*. [Applications of Mathematics **27**] Springer-Verlag, Berlin.

Sunder Sethuraman  
 400 Carver Hall  
 Department of Mathematics  
 Iowa State University  
 Ames, IA 50011  
 sethuram@iastate.edu

S.R.S. Varadhan  
 251 Mercer St.  
 Courant Institute  
 New York University  
 New York, NY 10012  
 varadhan@cims.nyu.edu